(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number : 0 290 190 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification : 09.10.91 Bulletin 91/41

(51) Int. Cl.[5] : G10L 5/06

(21) Application number : 88303794.7

(22) Date of filing : 27.04.88

(54) Pattern matching system.

(30) Priority : 30.04.87 JP 104632/87
01.05.87 JP 106420/87
01.05.87 JP 106421/87
06.05.87 JP 108815/87
06.05.87 JP 108818/87

(43) Date of publication of application : 09.11.88 Bulletin 88/45

(45) Publication of the grant of the patent : 09.10.91 Bulletin 91/41

(84) Designated Contracting States : GB NL SE

(56) References cited :
ICASSP 79, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING, Washington, 2nd-4th April 1979, pages 566-569, IEEE, New York, US; H. MATSUMOTO et al.: "Frequency warping for nonuniform talker normalization"
ICASSP 79, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING, Washington, 2nd-4th April 1979, pages 235-238, IEEE, New York, US; J. JASCHUL: "An approach to speaker normalization for automatic speech recognition"
ICASSP 83, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Boston, Massachusetts, 14th-16th April 1983, vol. 2, pages 769-772, IEEE, New York, US; R.J. GOLIBERSUCH: "Automatic prediction of linear frequency warp for speech recognition"
JOURNAL OF ACOUSTICAL SOCIETY OF AMERICA, vol. 68, no. 4, October 1980, pages 1071-1076, Acoustical Society of America, New York, US; H.W. STRUBE: "Linear prediction on a warped frequency scale"

(56) References cited :
SYSTEMS-COMPUTERS-CONTROLS, vol. 12, no. 1, January-February 1981, pages 28-37, Scripta Publishing Co., Silver Spring, Maryland, US; J. MIWA et al.: "A method of spoken word recognition using nonlinear spectral matching"
ICASSP 86, IEEE-IECEJ-ASJ INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Tokyo, 7th-11th April 1986, vol. 4 of 4, pages 2647-2650, IEEE, New York, US; N. HIGUCHI et al.: "Speaker-adaptation methods using selective linear prediction"
37TH IEEE VEHICULAR TECHNOLOGY CONFERENCE, Tampa, Florida, 1st-3th June 1987, pages 547-551, IEEE, New York, US; Y. TAKIZAWA et al.: "A new speaker-independent voice recognition scheme for voice dialling"

(73) Proprietor : Oki Electric Industry Company, Limited
7-12, Toranomon 1-chome Minato-ku
Tokyo 105 (JP)

(72) Inventor : Takizawa, Yumi c/o Oki Electric Industry Co., Ltd.
7-12, Toranomon 1-chome
Minatoku Tokyo (JP)
Inventor : Sato, Shinichi c/o Oki Electric Industry Co., Ltd.
7-12, Toranomon 1-chome
Minatoku Tokyo (JP)
Inventor : Fukasawa, Atsushi Oki Electric Industry Co., Ltd.
7-12, Toranomon 1-chome
Minatoku Tokyo (JP)
Inventor : Sato, Takuro c/o Oki Electric Industry Co., Ltd.
7-12, Toranomon 1-chome
Minatoku Tokyo (JP)
Inventor : Ando, Hiromi c/o Oki Electric Industry Co., Ltd.
7-12, Toranomon 1-chome
Minatoku Tokyo (JP)
Inventor : Suzuki, Yukio c/o Oki Electric Industry Co., Ltd.
7-12, Toranomon 1-chome
Minatoku Tokyo (JP)

(74) Representative : Read, Matthew Charles et al
Venner Shipley & Co. 368 City Road
London EC1V 2QA (GB)

EP 0 290 190 B1

Description

This invention relates to a system for pattern matching on the frequency axis, for use in speech recognition equipment.

Known pattern matching systems comprise input means to receive signals which are a function of an input speech signal, reference signal means for producing a reference signal representative of input speech, comparison means for making a comparison between the signals from the input means and the reference signals, and adaptive control means for adapting the signals used in the comparison in dependence upon the result thereof.

An adaptive system of this kind is described in ICASSP 79, IEEE International Conference on Acoustics, Speech & Signal Processing, Washington, 2nd-4th April 1979, pages 566-569, IEEE, New York, US ; H. Matsumoto et al : "Frequency warping for nonuniform talker normalisation". The reference discloses a non-linear spectral normalisation technique which eliminates inter-speaker differences from frequency band limited speech. A frequency normalised distance between a test and a reference spectrum is defined on the basis of minimum mean square difference calculations performed over a number of different frequency warping functions.

Also, a linear transformation of an autocorrelation function is known per se from J. Acoust Am Vol 68 (1980) Oct., No. 4 New York pages 1071-1075.

In practice, when speaker-independent speech recognition apparatus performs pattern matching, it normalises individual differences which are present on both the frequency axis and the time axis. An example of a prior-art system of pattern matching on these two axes is disclosed in the paper "Speaker-Independent Spoken Word Recognition Based on Time-Frequency-Intensity Warping of Spectra", Nakagawa et al, The Transactions of the Inst. of Electronics and Communication Engineers of Japan '81/2 Vol. J64-D No. 2. This method performs dynamic-programming matching (DPM) on the frequency axis and time axis of discrete output values from bandpass filters.

A problem with this pattern matching system is that since it uses features obtained from the output of a bank of bandpass filters, it has difficulty in handling the details of frequency patterns accurately. Another method uses the fast Fourier transform (FFT) instead of a filter bank, but this method has the problem of confusing voice pitch effects with vocal-tract effects, and can provide only inexact results. A difficulty associated with speech recognition is that speech signal contains rapid changes along frequency axis.

With a view to overcoming these problems and difficulties, the present invention provides a pattern matching system wherein the signals received by said input means comprises linear predictive analysis coefficients of the input speech signal ; and including transformation means including means for providing a series of outputs corresponding to the frequency transformation performed by a sequence of all-pass filters on a unit impulse, means for receiving the linear predictive analysis coefficients for use as respective tap coefficients for weighting said outputs, and means for combining the resulting weighted output signals to provide frequency transformed linear predictive analysis coefficients ; the reference signal means being operative to output linear predictive analysis coefficients corresponding to a frequency transformed reference signal derived from linear predictive analysis coefficient of a reference speech signal, and said adaptive control means including means for calculating a distance between the respective linear predictive analysis coefficients from the transformation means and the reference signal means, means for controlling the values of at least one of the frequency transformed signals such as to minimise said distance, and means responsive to said distance for generating a pattern matching residual on the frequency axis.

Thus, the present invention utilises linear predictive analysis coefficients, which are frequency transformed for use in the adaptive matching procedure.

In order that the invention may be more fully understood, embodiments thereof will now be described by way of example with reference to the accompanying drawings in which :

Figure 1 is a block diagram showing a pattern matching system of an embodiment of this invention ;

Figure 2 is a diagram of the internal structure of the spectrum matching (SPM) section of the embodiment of Figure 1 ;

Figure 3A and Figure 3B show the internal structure of the frequency transformation sections in Figure 2;

Figure 3C is a diagram combining Figure 3A and Figure 3B to show the frequency transformation sections;

Figure 4 is a block diagram showing an example of the parameter decision section 33 of Figure 2 ;

Figure 5A is a schematic diagram of the internal structure of a filter for finding the variations $g_{i(n)}$;

Figure 5B is a schematic diagram of a modification of the filter of Figure 5A ;

Figure 6 explains the pattern matching process performed in the DPM section ;

Figure 7 is a schematic diagram showing another example of the structure of the frequency transformation sections ;

Figure 8 is a block diagram explaining the principle of operation of another embodiment of this invention ;
Figure 9 is a schematic diagram of this embodiment ;
Figure 10 is a block diagram showing a modification of the frequency transformation circuit ;
Figure 11A and Figure 11B are characteristic curves showing experimental data for the frequency transformation circuit of Figure 10 ;
Figure 12 is a block diagram for explaining the principle of operation of another embodiment of the frequency transformation circuit ; and
Figure 13 is a block diagram showing the embodiment of Figure 12.

Figure 1 is a block diagram showing an embodiment of this invention. In this figure the numeral 1 denotes a memory for storing linear predictive coefficients obtained as features from an input speech signal (input pattern), 2 denotes a memory for temporarily storing linear predicitive coefficients obtained as features from a reference speech signal (standard pattern), 3 denotes a spectrum pattern matching (SPM) section that performs pattern matching on the frequency axis based on the signal output from the memories signals from these all-pass filters by tap coefficients, and an adder for adding the signals obtained from these means.

With this structure, the parameters defining the transfer characteristics of the all-pass filters are selected according to a desired purpose, and the tap coefficients are set in accordance with desired sample values depending on the desired purpose. In consequence, a transformation is effected between the frequency structure of an input signal and the frequency structure of a time series obtained as an output signal from a unit impulse input.

The frequency transformation circuit of the above structure has the ability to perform a variety of frequency structure transformations.

According to another aspect of the invention, there is provided a frequency transformation circuit which comprises storage means for storing parameters defining the designated transfer characteristics of a plurality of all-pass circuits, a parameter determining the size of a frequency transformation to be performed on an input signal, and constant values of a unit impulse series ; and multiply-add means for multiplying and adding these constant values with the input signal.

With the above structure, the constant values necessary for frequency transformation of the input signal are read from the storage means wherein they are stored. The multiply-add means performs a multiply-add operation on the input signal and corresponding constant values, as a result of which a frequency-transformed output signal is obtained.

The frequency transformation circuit of this configuration is simple in structure, and small-sized.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing a pattern matching system of an embodiment of this invention.
Fig. 2 is a diagram of the internal structure of the SPM section of the embodiment in Fig. 1.
Fig. 3A and Fig. 3B show the internal structure of the frequency transformation sections in Fig. 2.
Fig. 3C is a diagram combining Fig. 3A and Fig. 3B to show the frequency transformation sections.
Fig. 4 is a block diagram showing an example of the parameter decision section 33 of Fig. 2.
Fig. 5A is a shematic diagram of the internal structure of a filter for finding the variation $g_{l(n)}$.
Fig. 5B is a schematic diagram of a modification of the filter of Fig. 5A.
Fig. 6 explains the pattern matching process performed in the DPM section.
Fig. 7 is a schematic diagram showing another example of the structure of the frequency transformation sections.
Fig. 8 is a block diagram explaining the principle of operation of another embodiment of this invention.
Fig. 9 is a schematic diagram of this embodiment.
Fig. 10 is a block diagram showing a modification of the frequency transformation circuit.
Fig. 11A and Fig. 11B are characteristic curves showing experimental data the frequency transformation circuit of Fig. 10.
Fig. 12 is a block diagram for explaining the principle of operation of another embodiment of the frequency transformation circuit.
Fig. 13 is a block diagram showing the embodiment of Fig. 12.

DETAILED DESCRIPTION OF THE EMBODIMENTS

Fig. 1 is a block diagram showing an embodiment of this invention. In this figure the numeral 1 denotes a memory for storing linear predictive coefficients obtained as features from an input speech signal (input pattern), 2 is a memory for temporarily storing linear predictive coefficients obtained as features from a reference speech

signal (standard pattern), 3 is a spectrum pattern matching (SPM) section that performs pattern matching on the frequency axis based on the signals output from the memories 1 and 2, and 4 denotes a time pattern matching (DPM) section that performs pattern matching on the time axis, based on the output signals from the SPM section 3, by means of DP matching, and supplies the memories 1 and 2 with data m, k specifying time, so that the memories 1 and 2 produce the input data of the specific time.

The operation is explained next.

The memory 1 receives a set of linear predictive coefficients $a_{(m)} = \{a_0^{(m)}, a_1^{(m)}, \ldots, a_p^{(m)}\}$ (m = 1, 2,..., M ; time) of the input speech signal on which pattern matching is to be performed, and sends to the SPM section 3 the linear predictive coefficients $\bar{a}_{(m)}$ for the time specified by the signal m from the DPM section 4. Similarly, the memory 2 receives a set of linear predictive coefficients $\bar{a}_{(k)} = \{\bar{a}_0^{(k)}, \bar{a}_1^{(k)}, \ldots, \bar{a}_p^{(k)}\}$ (k = 0, 1,..., K ; time) of the reference speech signal and sends to the SPM section 3 the linear predictive coefficients $\bar{a}_{(k)}$ for the time specified by the signal k from the DPM section 4. The SPM section 3 performs pattern matching on the frequency axis with regard to the linear predictive coefficients $\bar{a}_{(m)}$ of the input speech signal and the linear predictive coefficients $\bar{a}_{(k)}$ of the reference speech signal. The residual de that remains is sent to the DPM section 4. Based on the residual de from the SPM section 3, the DPM section 4 performs pattern matching on the time axis by DP matching, and generates final output of the residual D of pattern matching on both the frequency axis and the time axis.

The internal structure of the SPM section 3 is shown in Fig. 2. The SPM section 3 comprises two frequency transformation sections 31 and 32, a distance calculation section 34, and a parameter decision section 33. The frequency transformation section 31 for the input signal receives the linear predictive coefficients $a_{(m)}$ and generates output of a set of linear predictive coefficients $\hat{a}_{(m)}$ resulting from frequency transformation according to a variable all pass filter parameter α received from the parameter decision section 33. The frequency transformation section 31 in Fig. 2 receives the all-pass filter parameter α from the parameter calculation section 33, continuously controls the frequency, and outputs $\hat{a}(m) = \hat{a}_0(m), \hat{a}_1(m), \ldots, \hat{a}_q(m)\}$. The parameters α determine the amount of frequency variation. The frequency transformation section 32 for the reference speech signal receives the reference speech linear predictive coefficients $\bar{a} = \{\bar{a}_0, \bar{a}_1, \ldots, \bar{a}_p\}$, performs a frequency transformation by the same process as in the frequency transformation section 31, and outputs the reference linear predictive coefficients

$$\tilde{a} = \{\tilde{a}_0, \tilde{a}_1, \ldots, \tilde{a}_q\}.$$

The frequency transformation section performs a standard frequency transformation. The parameter α_0 used for the transformation is fixed. In this embodiment, the standard transformation is a transformation to a psychological scale (mel frequencies).

Fig. 3A shows the structure of the frequency transformation section 31 of the input speech signal. The frequency transformation section 31 comprises all-pass filters 31z, multipliers 31c, and adders 31a. The frequency transformation section 31 is identical to an FIR filter except that delay elements in an FIR filter have been replaced by the all-pass filters 31z. The frequency transformation section receives linear predictive coefficients as tap coefficients. The impulse response of the all-pass filters 31z is the set of linear predictive coefficients $\hat{a}_{(m)} = \{\hat{a}_0^{(m)}, \hat{a}_1^{(m)}, \ldots, \hat{a}_p^{(m)}\}$ after frequency transformation. The frequency transformation section 32 of the reference speech signal is similar.

The general form of the transfer function of an all-pass filter is :

$$\Lambda(z) = \pm \prod_{i=1}^{N} (z^{-1} - \alpha_i)/(1 - \alpha_i z^{-1}) \tag{1}$$

Fig. 3B shows an example of the structure of an all-pass filter. The order of the filter shown is N, with each section comprising adders 301 and 303, multipliers 304 and 302 for the parameter α, and a delay element 305. In this embodiment the all-pass filters in each stage have the same order. If a first order all-pass filter is used, the relation between $a_{(m)}$ and $\hat{a}_{(m)}$ is :

$$\hat{a}_i^{(m)} = \frac{1}{2\pi j} \oint z^{i-1} \sum_{j=0}^{p} a_j^{(m)} \{(z^{-1} - \alpha)/(1 - \alpha z^{-1})\}^j dz \tag{2}$$

The frequency θ in the a $_{(m)}$ is transferred in $\hat{a}$ to :

$$\omega = \tan^{-1}\{(1-\alpha^2)\sin\theta/2 + (1+\alpha^2)\cos\theta\} \quad (3)$$

Fig. 3C is a diagram combining Fig. 3A and Fig. 3B to show the frequency transformation section 31.

The principle of operation of the frequency transformation section will be explained with reference to Fig. 3C.

The numeral 41 denotes an input terminal for the parameter α determining the amount of frequency transformation. Reference 42 denotes an input terminal for a digital unit impulse series {1, 0,..., 0}, 43-0 to 43-P are input terminals for the linear predictive coefficients of the input signal, 31c-0 to 31c-P are multipliers, 31z-1 to 31z-P are all-pass filters, 31a-1 to 31a-P are adders, and numeral 47 denotes an output terminal.

The all-pass filters 31z-1 to 31z-P comprise an adder 301, a delay element 305 a multiplier 302 a multiplier 304 and an adder 303.

The frequency transformation circuit with this configuration operates as follows. First a constant value α is received through the input terminal 41 and a unit impulse series (1, 0,..., 0) of P + 1 samples is received through the input terminal 42. At point 49-0, accordingly, the input is a unit impulse series, while at points 49-1 to 49-P the input is a series of constant values comprising the constant value α and a unit impulse series. The multipliers 31c-0 to 31c-P and the adders 31a-1 to 31a-P perform a multiply-add operation on the constant values at the points 49-0 to 49-P and the input signals received through the terminals 43-0 to 43-P to obtain the sum of the products of the constant values at the points 49-0 to 49-P and the input signals received at the input terminals 43-0 to 43-P and this sum becomes the output at the output terminal 47. The output is a time series representing the frequency transformed signal.

After frequency transformation, the distance calculation section 34 in Fig. 2 calculates the distance (difference) d between the two sets of the linear predictive coefficients $\hat{a}_{(m)}$ and $\tilde{a}_{(k)}$ by the following formula :

$$\begin{aligned} d &= \hat{a}_{(m)} - \tilde{a}_{(k)} \\ &= \{\hat{a}_0^{(m)} - \tilde{a}_0^{(k)},\ \hat{a}_1^{(m)} - \tilde{a}_1^{(k)},\ \ldots,\ \hat{a}_q^{(m)} - \tilde{a}_q^{(k)}\} \\ &= \{d_0,\ d_1,\ \ldots,\ d_q\} \end{aligned} \quad (4)$$

The distance d obtained as the result is sent to the parameter decision section 33.

The parameter decision section 33 continuously updates the parameter α using the nonlinear-parameter least-squares method to minimize the distance d. When the circuit shown in Fig. 3B is used as the all-pass filter, then if $\alpha_{(n)}$ is the all-pass filter parameter α after the n-th update and $d_{(n)} = \{d_{0(n)}, d_{1(n)},\ldots, d_{q(n)}\}$ is the distance d after the n-th update, then the parameter $\alpha_{(n+1)}$ after the next update is given by :

$$\alpha(n+1) = \alpha(n) - \delta\alpha(n)$$

$$\delta\alpha(n) = \sum_{i=0}^{q} d_i(n)\cdot g_i(n) \Big/ \sum_{i=0}^{q}\{g_i(n)\}^2 \quad (5)$$

The parameter decision section 33 receives the distance

$$d(n) = \{d_0(n), d_1(n),\ldots, d_i(n),\ldots, d_q(n)\}$$

and outputs the parameter α(n + 1) after the update. An example of the parameter determing section 33 is shown in Fig. 4. As illustrated, the correction amount δα(n) is calculated at a correction amount calculator 33i, from the output $g_{i(n)}$ of a gradient filter 33h and $d_{i(n)}$. The parameter α(n + 1) after the update is determined at an adder 33k, from the present parameter α(n) stored in a register 33j and the correction amount δα(n) from the correction amount calculator 33i.

The gradient filter 33h determines the variation $g_{i(n)}$ of $\hat{a}_{i(m)}$ with respect to α(n).

An example of a filter for determining variations $g_{i(n)}$ of a_i with respect to $\alpha(n)$ is shown in Fig. 5A. This filter comprises all-pass filters (A(z)) 33z (substituting delay elements in an FIR filter), multipliers 33c, adders 33a, and a differentiating filter (A'(z)) 33d. The variations $g_{i(n)}$ are obtained as a time series from the impulse response of the filters shown. In Fig. 5A the differentiating filter 33d is located on the output side, but it could also be located on the unit impulse input side as shown in Fig. 5B. That is, the differentiating filter 33d is placed in a prestage before the FIR filter 33f configured similarly to the frequency transformation circuit. The output of this differentiating filter 33d is a constant value corresponding to the input unit impulse and the input parameter α, and the variation g can be obtained by a simple multiply-add calculation using a ROM circuit. More specifically, the variations $g_{i(n)}$ are given by a time series according to the impulse response of the filter :

$$g_{i(n)} = \frac{1}{2\pi j}\oint z^{i-1}\sum_{j=1}^{p} a_j^{(m)}\, j \{(z^{-1}-\alpha_{(n)})/(1-\alpha_{(n)}z^{-1})\}^{j-1} (z^{-2}-1)/(1-\alpha_{(n)}z^{-1})^2\, dz \qquad (6)$$

The initial value of α is set equal to the fixed parameter α_0 of the frequency transformation section 32 for the reference speech signal.

When the difference between the parameter $\alpha_{(n)}$ and the updated parameter $\alpha_{(n+1)}$ is such that $|\{\alpha_{(n+1)} - \alpha(n)\}/\alpha_{(n)}| < \varepsilon$ (where ε is a threshold value), the distance calculation section 34 generates the output de according to the following formula :

$$de = \sum_{i=0}^{p} (d_{i(n+1)})^2 \qquad (7)$$

This de is the residual left after adaptive frequency control to minimize the distance of the input linear predictive coefficients a. That is, d_e is the residual resulting from pattern matching of the input signal and reference signal on the frequency axis.

Let de(m, k) be the residual resulting from pattern matching on the frequency axis performed by the SPM section 3 on the linear predictive coefficients $a_{(m)}$ of the input speech signal and the linear predictive coefficients $\bar{a}_{(k)}$ of the reference speech signal.

Fig. 6 shows how the DPM section 4 performs pattern matching on the time axis on the basis of this residual de(m, k). The values of de(m, k) are located at the grid points in Fig. 6. The DPM section 4 determines the path from the point (1, 1) to the point (M, K) that minimizes the sum of the de(m, k) :

$$\sum_{(m,k)=0,0}^{(M+1),(K+1)} de(m, k).$$

This sum is determined by a recursive formula, the partial sum D(m, k) being given by :

$$D(m, k) = \min \begin{cases} D(m, k-1) + de(m, k) \\ D(m-1, k-1) + 2de(m, k) \\ D(m-1, k) + de(m, k) \end{cases} \qquad (8)$$

where D(0, 0) = 0, D(m, 0) = ∞ (m $\neq$ 0), D(0, k) = ∞ (k $\neq$ 0), de(M + 1, K + 1) = 0. The final pattern matching residual De is :

$$De = D(M+1, K+1)/(M+K) \qquad (9)$$

The DPM section 4 generates this residual De as its output.

The pattern matching method of the DPM section 4 described above does not impose any restrictions on the path, but it is also possible to determine D(m, k) as :

$$D(m, k) = \min \begin{cases} D(m-1, k-2) + 3de(m, k) \\ D(m-1, k-1) + 2de(m, k) \\ D(m-2, k-1) + 3de(m, k) \end{cases} \quad (10)$$

where D(0, 0) = 0, $D(m, k) = \infty$ when $m \leq 0$ and $k \neq 0$, $D(m, k) = \infty$ when $m \neq 0$ and $k \leq 0$, and de(M + 1, K + 1) = 0, and obtain the pattern matching residual

$$De = D(M + 1, K + 1)/(M + K).$$

The embodiment described above uses a method of controlling the parameter α of the SPM section 3 in which all the initial values are $\alpha = \alpha_0$ (constant), but it is possible to hasten the convergence of the parameter α by using as the initial value the converged value of the previous stage.

For simplicity, in the frequency transformation sections 31 and 32 of the SPM, the all-pass filters A(z) (31z) in all stages had the same parameter α, but it is also possible to use an all-pass filter (31'z) having a differing configuration (parameter value or order) in each stage. Fig. 7 shows an example of such a configuration.

Similarly, it should be clear that different conditions can be used as the DPM path restriction conditions.

In addition, although linear predictive coefficients were used as features in the foregoing description, linear predictive cepstral coefficients, linear predictive melcepstral coefficients, or other coefficients obtained by the linear predictive method could be used instead. The term "linear predictive analysis coefficient(s)" as used in the appended claims should therefore be construed to cover such alternatives.

Furthermore, in the preceding embodiment the parameter α of the frequency transformation section 31 of the input speech signal (input pattern), which is the delay (frequency transformation amount), was adaptively controlled, but it is also possible to adaptively control the delay of the frequency transformation section 32 of the reference pattern, or to control the delays of both frequency transformation sections adaptively, although that would require an increased amount of computation.

In the above configurations, pattern matching functions are provided for both the frequency axis and the time axis and pattern matching is performed simultaneously on these two axes using linear predictive analysis coefficients, so the matching process is accurate and simple. As a pattern matching method for use in speaker-independent speech recognition apparatus, this method can be expected to improve the recognition rate.

Fig. 8 and Fig. 9 show another embodiment of this invention. Fig. 8 is a block diagram explaining the principle of operation of this embodiment. Fig. 9 is a schematic diagram of a pattern matching circuit illustrating an embodiment of this invention.

First the principle of operation of this embodiment will be explained with reference to Fig. 8. In Fig. 8 reference numeral 51 denotes an input terminal for the linear predictive coefficients of an input signal (input pattern), numeral 52 denotes a frequency transformation and distance calculation circuit, numeral 53 denotes an α updating variation detection circuit, numeral 54 denotes a residual detection circuit, numeral 55 denotes an updating circuit, numeral 56 denotes a decision circuit (DEC), numeral 57 denotes a switch (SW), numeral 58 denotes an output terminal, numeral 59 denotes an input terminal for input of linear predictive coefficients of a reference pattern (standard pattern), and numeral 60 denotes a frequency transformation circuit. The time series $a = \{a_0, a_1, ..., a_p\}$ of the linear predictive coefficients of the input signal through the input terminal 51 are input as tap coefficients to and frequency-transformed by the frequency transformation circuit 52 which may be similar to that shown in Fig. 3A through Fig. 3C, which acts on a unit impulse series. The frequency transformation circuit is similar to an FIR filter, but its delay elements have been replaced by all-pass filters. Thus, the frequency transformation circuit of the frequency transformation and distance calculation circuit 52 performs frequency transformation corresponding to a parameter α that defines the all-pass filters of an FIR filter to produce an output impulse series $\hat{a}$.

The time series $\bar{a} = \{\bar{a}_0, \bar{a}_1, ..., \bar{a}_p\}$ of linear predictive coefficients of the reference signal is subjected only to a standard frequency transformation in the frequency transformation circuit 60, obtaining $\tilde{a} = \{\tilde{a}_0, \tilde{a}_1, ..., \tilde{a}_q\}$. In this case the parameter defining the all-pass filters has the fixed value α_0.

Next, the distance calculation circuit in the frequency transformation and distance calculation circuit 52 gen-

erates the distance d between the frequency-transformed linear predictive coefficients $\hat{a}$ and $\tilde{a}$, where

$$d = \hat{a} - \tilde{a} = \{d_0, d_1, \ldots, d_q\} = \{\hat{a}_0 - \tilde{a}_0, \hat{a}_1 - \tilde{a}_1, \ldots, \hat{a}_q - \tilde{a}_q\}.$$

At this time, an impulse response g expressing the variation in $\hat{a}$ when the parameter α is moved is obtained from the α updating variation detection circuit 53, which comprises an FIR filter similar to the frequency transformation circuit.

The parameter α is continuously updated so as to minimize the distance d. If $\alpha(n)$ is the parameter α after the n-th update and $d(n) = \{d_0^{(n)}, d_1^{(n)}, \ldots, d_q^{(n)}\}$ is the distance d after the n-th update, then the parameter $\alpha(n + 1)$ after the next update is given by :

$$\alpha(n+1) = \alpha(n) - \sum_{i=0}^{q} d_i(n) \cdot g_i(n) / \sum_{i=0}^{q} \{g_i(n)\}^2$$

This calculation is performed by the α updating circuit 55. The initial value of α is the value α_0 used in the frequency transformation of the linear predictive coefficients $\bar{a}$ of the reference signal.

When the decision circuit 56 obtains from the parameter $\alpha(n)$ and the updated parameter $\alpha(n + 1)$ the result $|\{\alpha(n + 1) - \alpha(n)\}/\alpha(n)| < \varepsilon$ (where ε is a threshold value), the residual detection circuit 54 outputs the result of the following calculation through the switch 57 :

$$d_e = \sum_{i=0}^{q} \{d_{i(n+1)}\}^2$$

This d_e is the residual left when frequency control is adaptively performed to minimize the distance between the linear predictive coefficients a of the input signal and the reference linear predictive coefficients $\bar{a}$; that is, it is the residual left after pattern matching on the frequency axis. The frequency transformation circuit in the frequency transformation and distance calculation circuit 52 can be the one identical to that shown in Fig. 3A to Fig. 3C.

An explanation will now be given of the operation of the circuit in Fig. 9, which embodies the principles of operation just described for pattern matching on the frequency axis. Component elements in Fig. 9 that are identical to elements in Fig. 8 are indicated by the same reference numerals. The memory (MEM) 60a holds precalculated values of the linear predictive coefficients $\tilde{a}$ obtained by frequency transformation of the linear predictive coefficients $\bar{a}$ of the reference signal and provides output of $-\tilde{a}$ in order to calculate a distance. The frequency transformation and distance calculation circuit 52 comprises a parameter memory (MEM) 52a, a multiplier (MPY) 52b, an adder (ADD) 52c, and a register (REG) 52d. The α updating variation detection circuit 53 comprises a parameter memory (MEM) 53a, a multiplier (MPY) 53b, an adder (ADD) 53c, and a register (REG) 53d.

The residual detection circuit 54 comprises a multiplier (MPY) 54a, an adder (ADD) 54b, and a register (REG) 54c. The α updating circuit 55 comprises multipliers (MPY) 55a and 55d, adders (ADD) 55b, 55e, and 55h, registers (REG) 55c, 55f, and 55i, and a divider (DIV) 55g.

The operation of this embodiment is described next.

First the time series a of linear predictive coefficients of the input signal is input from the input terminal 51 to the frequency transformation and distance calculation circuit 52. This circuit 52 uses the multiplier 52b, the adder 52c, and the register 52d to multiply and add the input linear predictive coefficients with the series of constant values output from the memory 52a (which consist of the parameter α and an impulse series as described earlier). If the frequency-transformed reference data -a stored in the memory 60a are placed in the register 52d as initial values, the result of the multiply-add calculation described above is the distance d.

At the same time, the linear predictive coefficients are input by the α updating variation detection circuit 53. This circuit 53 uses the multiplier 53b, the adder 53c, and the register 53d to multiply and add the series of constant values output from the memory 53a and the input linear predictive coefficient series, yielding as output the variations g. The initial value of the register 53d in this operation is "0."

Next, the residual detection circuit 54 uses the multiplier 54a to square the distance time series d received from the α updating variation circuit 53, and uses the adder 54b and register 54c (which is initialized to "0") to

add the results and obtain the residual d_*. The true matching residual is not obtained, however, unless the values of α converge. The control for this purpose will be described later.

Next the α updating circuit 55 uses the multiplier 55a, the adder 55b, and the register 55c to multiply and add the distance time series d with the output g of the α updating variation detection circuit 53 to obtain $\Sigma d \cdot g$. At the same time, the multiplier 55d squares the output of the updating variation detection circuit 53, and the adder 55e and register 55f obtain the sum Σg^2. Next the divider 55g performs the division $\Sigma d \cdot g/\Sigma g^2$, and the adder 55h adds the previous value of α in the register 55i (at the start time this is the initial value) to obtain a new value of α.

The decision circuit 52c decides whether the value of α obtained in this way has converged or not. If it decides that the value has not converged, the updated value of α is used to change the values in the memory 52a and the memory 53d. If the range of is limited, the memory 52a and the memory 53d can be implemented as ROM circuits containing, in each memory, a calculated value for each value of α that can be selected.

The above-described operation is repeated until the value of α converges. When the decision circuit 56 decides that the value has converged, it closes the output control switch 57, and the output d_* of the residual detection circuit 54 calculated with the value of α determined at this time becomes the output at the output terminal 58 representing the pattern matching residual on the frequency axis.

In the above configuration, a first and a second series of constant values corresponding to parameter values determining frequency transformation quantities are held in memory means, multiply-add operations are performed on these series of constant values and linear predictive coefficients representing features to calculate frequency-transformed linear predictive coefficients, distances, and variations, and parameter matching on the frequency axis is performed with adaptive control of the parameters, so the circuit can be compact and simple in configuration, and precise pattern matching on the frequency axis can be performed. It is thus possible to provide inexpensive speech recognition equipment with an improved rate of recognition.

A modification of the frequency transformation circuit incorporated in the above-described pattern matching system is shown in Fig. 10.

In Fig. 10 the numerals 101-0, 101-1,..., 101-P, 102-1, 102-2,..., 102-N, and 103 denote terminals, 104 is an oscillator, 105-1, 105-2,..., 105-P are Nth-order all-pass-filters (APFs), 106-0, 106-1, 106-2,..., 106-P are amplifiers, and 107 is an adder. The Nth-order APFs 105-1, 105-2,..., 105-P all have the same transfer function:

$$A \prod_{k=1}^{N} \frac{Z^{-1}-\alpha_k}{1-\alpha_k Z^{-1}} = A \cdot \frac{Z^{-1}-\alpha_1}{1-\alpha_1 Z^{-1}} \cdot \frac{Z^{-1}-\alpha_2}{1-\alpha_2 Z^{-1}} \cdots\cdots \frac{Z^{-1}-\alpha_N}{1-\alpha_N Z^{-1}}$$

where the α_k are constants such that $|\alpha_k| < 1$ (k = 1,..., N) and A is the constant 1 or – 1. Values in the range from – 1 to 1, corresponding to the desired frequency transformation characteristic, are input at the terminals 102-1,..., 102-N, and the value 1 or – 1 is input at the terminal 103. The values from the terminals 102-1,..., 102-N are furnished to all the Nth-order APFs 105-1, 105-2,..., 105-P to set the values $\alpha_1, \alpha_2, \ldots, \alpha_N$ of the transfer function of the APFs 105-1, 105-2,..., 105-P. The input value at the terminal 103 is also furnished to all the Nth-order APFs 105-1, 105-2,..., 105-P to set the value A is the transfer function of the Nth-order APFs 105-1, 105-2, ..., 105-P. The terminals 101-0, 101-1, 101-2,..., 101-P receive the first sample data, the second sample data, the third sample data,..., the (P + 1)-th sample data of the digital signal to be frequency-transformed. The data from the terminals 101-0, 101-1, 101-2,..., 101-P are supplied to the amplifiers 106-0, 106-1, 106-2,..., 106-P to set the gain of the amplifiers 106-0, 106-1, 106-2,..., 106-P. Next an output signal from the oscillator 104, which generates a digital unit impulse signal, is applied to the amplifier 106-0 and the Nth-order APF 105-1. P-1 Nth-order APFs, namely the Nth-order APFs 105-2, 105-3,..., 105-P, are connected in series to the output of the Nth-order APF 105-1. The outputs from the Nth-order APFs also become the inputs to the amplifiers 106-1, 106-2,..., 106-P. Accordingly, the amplifier 106-0 amplifies the output signal from the oscillator 104 with a gain set by the data from the terminal 101-0. The amplifiers 106-1, 106-2,..., 106-P amplify the output signals from the Nth-order APFs 105-1, 105-2,..., 105-P with gains set by the data from the terminals 101-1, 101-2,..., 101-P. The output signals from the amplifiers 106-0, 106-1, 106-2,..., 106-P are received by the adder 107, which adds all the output signals from the amplifiers 106-0, 106-1, 106-2,..., 106-P. The output from the adder 107 is a signal with a transformed frequency structure, derived from a digital signal having as its first sample, second sample, third sample,..., (P + 1)-th sample the data from the terminals 101-0, 101-1, 101-2,..., 101-P by a frequency structure transformation with a characteristic defined by values of $\alpha_1, \alpha_2, \ldots, \alpha_N$ and A in the formula for the transfer function of the APFs 105-1, 105-2,..., 105-P set by the values at the terminals 102-1, 102-2,..., 102-N, 103.

Fig. 11A and Fig. 11B are characteristic curves showing experimental data obtained with the circuit of the above configuration. In both diagrams frequency (Hz) is shown on the horizontal axis and power (dB) on the vertical axis. First-order APFs with the transfer function $A(Z^{-1} - \alpha)/(1 - \alpha Z^{-1})$ were employed as the Nth-order APFs. The input signal consisted of 129 samples of a 1001Hz sine wave sampled at a rate of 8kHz. As the parameters A and α that determined the frequency structure transformation characteristic, a value of 1 was used for A and a value of -0.5 for α. As can be seen by examining the transformation from Fig. 11A, which shows the power spectrum of the untransformed signal, to Fig. 11B, which shows the power spectrum of the transformed signal, it is possible to transform the frequency characteristic without changing the power.

In this embodiment, once the values from the terminals 102-1, 102-2,..., 102-N, 103 are set they remain fixed during circuit operation, but it would be useful for a plurality of adaptive purposes to vary these values over time. Also, although the transfer functions of all the Nth-order APFs 105-1, 105-2,..., 105-P were identical, it would be useful for a plurality of adaptive purposes to vary the transfer functions of the Nth-order APFs 105-1, 105-2,..., 105-P with respect to each other.

As explained above, in the above configuration, there are many parameters for determining the frequency structure transformation characteristic, so a variety of frequency structure transformations can be effected. If provided with speech signal input, the circuit of the above configuration can also be used in speech signal processing apparatus. If provided with the impulse response of a filter as input, this circuit can further be used in a filter characteristic transformer, and if provided with speech recognition features as input it can be used in a speech feature transformer.

Another modification of the frequency transformation circuit is shown in Fig. 12 and Fig. 13.

It is seen from Fig. 3C, that the values at the points 49-0, 49-1, 49-2,..., 49-P are constants. It follows that the all-pass circuits 31z-1, 31z-2,..., 31z-P in Fig. 3C can be configured as in Fig. 12 using ROMs 221-1,..., 221-P wherein the constant values are stored. In other words, the time-series output of the frequency-transformed signal can be obtained from the output terminal 223 in Fig. 12 by means of a multiply-add operation performed on constant values from the ROMs 221-1,..., 221-P and the input signals received from the terminals 222-0, 222-1,..., 222-P.

It follows that the circuit of this embodiment can be configured as shown in Fig. 13. In Fig. 13 the numeral 201 denotes a ROM, 202 is a RAM, 203 is a parallel multiplier, 204 is an adder, 205 is a register (REG), and 47 is an output terminal. The ROM 201 corresponds to the ROMs 21-1,..., 21-P in Fig. 12. Locations $X_{0\text{-}P}$ in the ROM 201 contain a sequence of values corresponding to the constant values at the points 49-0, 49-1, 49-2,..., 49-P in Fig. 3C when the unit impulse series value is 1. Locations $Y_{0\text{-}P}$ in the ROM 201 contain a sequence of values corresponding to the constant values at the points 49-0, 49-1, 49-2,..., 49-P in Fig. 3C when the next unit impulse series value is 0. Locations $Z_{0\text{-}P}$ in the ROM 201 contain a sequence of values corresponding to the constant values at the points 49-0, 49-1, 49-2,..., 49-P in Fig. 3C when the P + 1-th sample of the unit impulse series value is 0. The RAM 202 contains the input signal series from the terminals 43-0, 43-1,..., 43-P.

Next the operation of the above-described circuit will be explained with reference to Fig. 13.

First the multiplier 203 multiplies X_0 in the ROM 201 with H in the RAM 202. The result is added by the adder 204 to the initial 0 output from the register 205 and the sum is placed in the register 205. Next the multiplier 203 multiplies X_1 in the ROM 201 with I in the RAM 202. The result is added by the adder 204 to the previous result in the register 205 and the sum is placed in the register 205. This process is repeated through $X_{0\text{-}P}$ in the ROM 201 and L in the RAM 202, and a time series output No. 0 is obtained from the output terminal 47. Next the same process is performed on the $Y_{0\text{-}P}$ in the ROM 201 and H, I,..., L in the RAM 202, obtaining the next time series output. If this process is repeated through $Z_{0\text{-}P}$ in the ROM 201 and H, I,..., L in the RAM 202, the entire frequency-transformed output is obtained from the output terminal 47.

The above-described frequency transformation circuit can be made small in scale and simple in circuit configuration.

Claims

1. A system for pattern matching on the frequency axis, comprising input means (1, 31) to receive signals which are a function of an input speech signal, reference signal means (2, 32) for providing a reference signal representative of input speech, comparison means (34) for making a comparison between the signals from the input means and the reference signals, and adaptive control means (33) for adapting the signals used in the comparison in dependence upon the result thereof, **characterised by**

   (a) the signals received by said input means comprising linear predictive analysis coefficients (43) of the input speech signal ;

   (b) transformation means (31) including means for providing a series of outputs corresponding to the fre-

quency transformation performed by a sequence of all-pass filters (31z-1, 31z-2...) on a unit impulse (42), means (31c-D, 31c-1, 31c-2...) for receiving the linear predictive analysis coefficients (43-1, 43-2) for use as respective tap coefficients for weighting said outputs, and means (31a-1, 31a-2, 31a-3...) for combining the resulting weighted output signals to provide frequency transformed linear predictive analysis coefficients ;

(c) the reference signal means (32) being operative to output linear predictive analysis coefficients corresponding to a frequency transformed reference signal derived from linear predictive analysis coefficient of a reference speech signal, and

(d) said adaptive control means (33) including means (34) for calculating a distance between the respective linear predictive analysis coefficients from the transformation means (31) and the reference signal means (32), means (41) for controlling the values of at least one of the frequency transformed signals such as to minimise said distance, and means responsive to said distance for generating a pattern matching residual (de) on the frequency axis.

2. A frequency transformation system according to claim 1, wherein said transformation means (31) comprises a plurality of all-pass filters (31z-1, 31z-2...) connected in series to provide said series of outputs, a first one of the all-pass filters having means (42) for receiving a unit impulse ; multiplying means (31c-D, 31c-1, 31c-2) for multiplying the outputs from the all-pass filters by respective ones of said tap coefficients (43-1, 43-2...) and an adder (31a-1, 31a-2...) for adding together signals obtained from said multiplying means to produce said frequency transformed linear predictive analysis coefficients ; and said adaptive control means (33) includes means (41) for controlling the transfer function characteristics of the all-pass filters.

3. A frequency transformation system according to claim 2 wherein the transfer functions of said all-pass filters are all the same.

4. A frequency transformation system according to claim 2 wherein the transfer functions of said all-pass filters differ from each other.

5. A frequency transformation system according to claim 2 wherein the transfer function of said all-pass filters are varied over time.

6. A frequency transformation system according to any preceding claim wherein said reference signal means (32) includes a reference filter including a sequence all-pass filters for performing a frequency transformation on an input unit impulse signal, means to receive linear predictive analysis coefficients of a reference speech signal as tap coefficients for the reference filter, said reference filter outputting said linear predictive analysis coeficient corresponding to the frequency transformed reference speech signal.

7. A frequency transformation system according to claim 6 wherein said adaptive control means (33) is operative to alter the filter characteristics of the reference filter.

8. A frequency transformation system according to claim 1 wherein said transformation means includes a plurality of memory means (221-1, 221-2...) for outputting said outputs corresponding to the frequency transformation performed by a sequence of all-pass filters on an input signal, multiplying means for multiplying said memory outputs by the linear predictive analysis coefficients (43-0, 43-1...) for the input speech signal, and adding means for adding the resultant signals produced by the multiplying means and thereby producing said frequency transformed linear predictive analysis coefficients (47) for the input speech signal.

9. A system according to claim 8, wherein the linear predictive analysis coefficients are cepstral coefficients.

10. A system according to any preceding claim, including dynamic pattern matching means (4) responsive to said pattern matching residual (de) for performing pattern matching on the time axis.

Patentansprüche

1. Anordnung zum Vergleichen von Mustern auf der Frequenzachse mit einer Eingangsvorrichtung (1, 31), um Signale, die eine Funktion eines Eingangssprachsignals sind, zu empfangen, einer Bezugssignalvorrichtung (2, 32), um ein Bezugssignal zu liefern, das Eingangssprache vertritt, einer Vergleichsvorrichtung (34), um einen Vergleich zwischen den Signalen von der Eingangsvorrichtung und den Bezugssignalen zu ziehen, und einer anpassungsfähigen Steuerungsvorrichtung (33), um die Signale, die in dem Vergleich in Abhängigkeit von dessen Ergebnis benutzt werden, anzupassen, gekennzeichnet durch

(a) die von der Eingangsvorrichtung empfangenen Signale mit linearen voraussagenden Auswertungskoeffizienten (43) des Eingangssprachsignals ;

(b) eine Umwandlungsvorrichtung (31), die eine Vorrichtung einschliesst, um eine Reihe von Ausgängen zu liefern, die der Frequenzumwandlung entspricht, welche von einer Folge von Allpassfiltern (31z-1, 31z-2...) auf einem Einheitsimpuls (42) durchgeführt wird, eine Vorrichtung (31c-D, 31c-1, 31c-2...), um die linearen voraussagenden Auswertungskoeffizienten (43-1, 43-2) zum Gebrauch als entsprechende

Abgriffskoeffizienten zu empfangen, um die Ausgänge zu bewerten, und eine Vorrichtung (31a-1, 31a-2, 31a-3...), um die sich ergebenden bewerteten Ausgangssignale zu kombinieren, um durch Frequenz umgewandelte lineare voraussagende Auswertungskoeffizienten zu liefern ;

(c) die Bezugssignalvorrichtung (32), betriebsfähig, um lineare voraussagende Auswertungskoeffizienten auszugeben, die einem durch Frequenz umgewandelten Bezugssignal entsprechen, das von dem linearen voraussagenden Auswertungskoeffizienten eines Bezugssprachsignals abgeleitet ist, und

(d) die anpassungsfähige Steuerungsvorrichtung (33) mit einer Vorrichtung (34) zum Errechnen einer Entfernung zwischen den entsprechenden linearen voraussagenden Auswertungskoeffizienten von der Umwandlungsvorrichtung (31) und der Bezugssignalvorrichtung (32), einer Vorrichtung (41) zum Steuern der Werte wenigstens eines der durch Frequenz umgewandelten Signale, um die Entfernung auf ein Mindestmass herabzusetzen, und einer auf die Entfernung ansprechende Vorrichtung, um einen Mustervergleichsrestwert (de) auf der Frequenzachse zu erzeugen.

2. Frequenzumwandlungsanordnung nach Anspruch 1, worin die Umwandlungsvorrichtung (31) eine Vielzahl von in Reihe geschalteten Allpassfiltern (31z-1, 31z-2...) enthält, um die Reihe von Ausgängen zu liefern, wobei ein erster der Allpassfilter eine Vorrichtung (42) hat, um einen Einheitsimpuls zu empfangen ; eine Multiplikationsvorrichtung (31c-D, 31c-1, 31c-2), um die Ausgänge von den Allpassfiltern mit entsprechenden der Abgriffskoeffizienten (43-1, 43-2...) zu multiplizieren, und einen Addierer (31a-1, 31a-2...), um Signale, die von der Multiplikationsvorrichtung erhalten werden, zu addieren, um die durch Frequenz umgewandelten linearen voraussagenden Auswertungskoeffizienten herzustellen ; und die anpassungsfähige Steuerungsvorrichtung (33) eine Vorrichtung (41) zum Steuern der Übertragungsfunktionskennzeichen der Allpassfilter enthält.

3. Frequenzumwandlungsanordnung nach Anspruch 2, worin die Übertragungsfunktionen der Allpassfilter sämtlich gleich sind.

4. Frequenzumwandlungsanordnung nach Anspruch 2, worin die Übertragungsfunktionen der Allpassfilter voneinander verschieden sind.

5. Frequenzumwandlungsanordnung nach Anspruch 2, worin die Übertragungsfunktion der Allpassfilter mit der Zeit geändert werden.

6. Frequenzumwandlungsanordnung nach einem der vorhergehenden Ansprüche, worin die Bezugssignalvorrichtung (32) einen Bezugsfilter mit einer Folge von Allpassfiltern einschliesst, um eine Frequenzumwandlung auf einem Eingangseinheitsimpulssignal durchzuführen, eine Vorrichtung, um lineare voraussagende Auswertungskoeffizienten eines Bezugssprachsignals als Abgriffskoeffizienten für den Bezugsfilter zu empfangen, wobei der Bezugsfilter den linearen voraussagenden Auswertungskoeffizienten ausgibt, der dem durch Frequenz umgewandelten Bezugssprachsignal entspricht.

7. Frequenzumwandlungsanordnung nach Anspruch 6, worin die anpassungsfähige Steuerungsvorrichtung (33) betriebsfähig ist, um die Filterkennzeichen des Bezugsfilters zu ändern.

8. Frequenzumwandlungsanordnung nach Anspruch 1, worin die Umwandlungsvorrichtung eine Vielzahl von Speichervorrichtungen (221-1, 221-2...) einschliesst, um die Ausgänge, die der Frequenzumwandlung entsprechen, welche von einer Folge von Allpassfiltern auf ein Eingangssignal durchgefahrt werden, auszugeben, eine Multiplikationsvorrichtung, um die Speicherausgänge mit den linearen voraussagenden Auswertungskoeffizienten (43-0, 43-1...) für das Eingangssprachsignal zu multiplizieren, und eine Additionsvorrichtung, um die sich ergebenden, von der Multiplikationsvorrichtung hergestellten Signale zu addieren, und dadurch die durch Frequenz umgewandelten linearen voraussagenden Auswertungskoeffizienten (47) für das Eingangssprachsignal herzustellen.

9. Anordnung nach Anspruch 8, worin die linearen voraussagenden Auswertungskoeffizienten Cepstralkoeffizienten sind.

10. Anordnung nach einem der vorhergehenden Ansprüche, die eine dynamische, auf den Mustervergleichsrestwert (de) ansprechende Vorrichtung zum Vergleichen von Mustern (4) einschliesst, um Vergleichen von Mustern auf der Zeitachse durchzuführen.

Revendications

1. Système d'appariage de modèle dans l'axe de fréquence, comportant des moyens d'entrée (1, 31) recevant des signaux fonction d'un signal vocal d'entrée, des moyens (2, 32) apportant un signal de référence représentatif d'entrée vocale, des moyens (34) de comparaison permettant la comparaison entre les signaux venant des moyens d'entrée et de référence, et des moyens de contrôle adaptatif (33) pour adapter les signaux exploités dans la comparaison en fonction du résultat, **caractérisé par,**

(a) les signaux reçus desdits moyens comportant des coefficients linéaires d'analyse prévisionnelle (43) du signal vocal d'entrée ;

(b) des moyens de transformation (31) comportant une série de sorties correspondant à la transformation de fréquence effectuée par une séquence de filtres passe-tout (31z..1, 31z-2..) sur une impulsion unitaire (42), des moyens de réception (31c-D, 31c-1, 31c-2..) de coefficients linéaires d'analyse prévisionnelle (43-1, 43-2) servant respectivement comme coefficients de captage pour la pondération desdites sorties, et des moyens (31a-1, 31a-2, 31a-3...) de combinaison des signaux pondérés qui en résultent pour obtenir des coefficients linéaires d'analyse prévisionnelle de fréquence transformée ;

(c) les moyens de signal de référence (32) étant fonctionnels pour la sortie de coefficients linéaires d'analyse prévisionnelles correspondant à un signal de référence de fréquence transformée dérivé du coefficient linéaire d'analyse prévisionnelle d'un signal vocal de référence, et

(d) lesdits moyens régulateurs d'adaptation (33) comportant des moyens (34) de calcul d'une distance entre les coefficients respectifs linéaires d'analyse prévisionnelle des moyens de transformation (31) et les moyens de signal de référence (32), des moyens régulateurs (41) de valeurs d'un minimum d'un signal de fréquence transformée de manière à minimiser ladite distance, et des moyens répondant à ladite distance pour la génération d'un résiduel (de) d'apparlage de modèle sur l'axe de fréquence.

2. Système de transformation de fréquence selon la revendication 1, dont lesdits moyens de transformation (31) prévoient une série de filtres passe-tout (31z-1, 31z-2..) raccordés en série pour assurer ladite série de sorties, un premier filtre passe-tout ayant des moyens (42) de réception d'impulsion unitaire ; des moyens multiplicateurs pour multiplier les sorties desdits filtres passe-tout par les sorties respectives desdits coefficients de captage (43-1, 43-2..) et un cumulateur (31a-, 31a-2..) assurant le cumul des signaux obtenus desdits moyens multiplicateurs pour obtenir lesdits coefficients linéaires d'analyse prévisionnelle en fréquence transformée ; et dont lesdits moyens régulateurs adaptateurs (33) prévoient des moyens régulateurs (41) des caractéristiques de fonction de transfert des filtres passe-tout.

3. Système de transformation de fréquence selon la revendication 2 dont les fonctions de transfert desdits filtres passe-tout sont toutes identiques.

4. Système de transformation de fréquence selon la revendication 2, dont les fonctions de transfert desdits filtres passe-tout varient l'une par rapport à l'autre.

5. Système de transformation de fréquence selon la revendication 2, dont les fonctions de transfert desdits filtres passe-tout varient dans le temps.

6. Système de transformation de fréquence selon l'une des revendications antérieures, dont lesdits moyens de signal de référence (32) comportent un filtre de référence prévoyant une série de filtres passe-tout pour effectuer la transformation de fréquence de signal d'entrée d'impulsion unitaire, des moyens de réception de coefficients linéaires d'analyse prévisionnelle d'un signal vocal de référence comme coefficients de captage du filtre de référence, ledit filtre de référence transmettant ledit coefficient linéaire d'analyse prévisionnelle correspondant au signal vocal de référence en fréquence transformée.

7. Système de transformation de fréquence selon la revendication 6, dont les moyens régulateurs-adaptateurs (33) sont fonctionnels pour modifier les caractéristiques de filtre de référence.

8. Système de transformation de fréquence selon la revendication 1, dont lesdits moyens de transformation prévoient une série de mémoires (221-1, 221-2..) pour la transmission desdites sorties correspondant à la transformation de fréquence effectuée par une série de filtres passe-tout sur un signal d'entrée, des moyens multiplicateurs pour multiplier lesdites sorties de mémoire par les coefficients linéaires d'analyse prévisionnelle (43-0, 43-1) pour le signal vocal d'entrée, et des moyens de cumul pour ajouter les signaux qui en résultent produits par les moyens multiplicateurs, assurant ainsi les coefficients linéaires d'analyse prévisionnelle (47) pour le signal vocal d'entrée.

9. Système selon la revendication 8, selon lequel les coefficients linéaires d'analyse prévisionnelle sont des coefficients de nature cepstrale.

10. Système selon l'une ou l'autre des revendications antérieures, comportant des moyens d'appariage de modèle dynamique (4) répondant audit résiduel (de) d'appariage de modèle pour assurer l'appariage de modèle dans l'axe du temps.

FIG.1

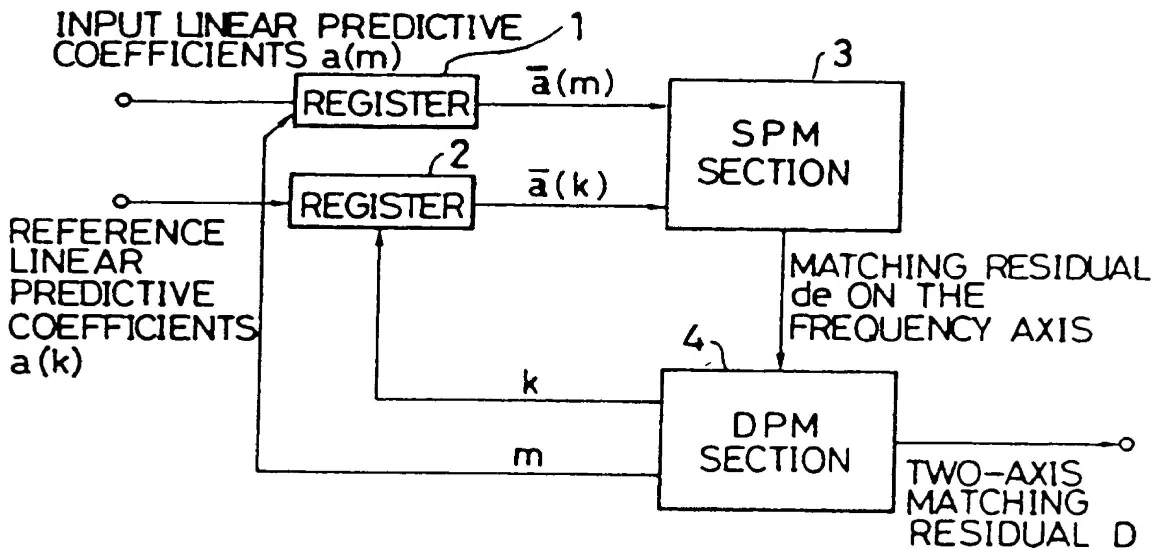

INPUT LINEAR PREDICTIVE COEFFICIENTS a(m)
1
REGISTER
$\bar{a}(m)$
3
SPM SECTION
2
REGISTER
$\bar{a}(k)$
REFERENCE LINEAR PREDICTIVE COEFFICIENTS a(k)
MATCHING RESIDUAL de ON THE FREQUENCY AXIS
4
k
DPM SECTION
m
TWO-AXIS MATCHING RESIDUAL D

FIG.2

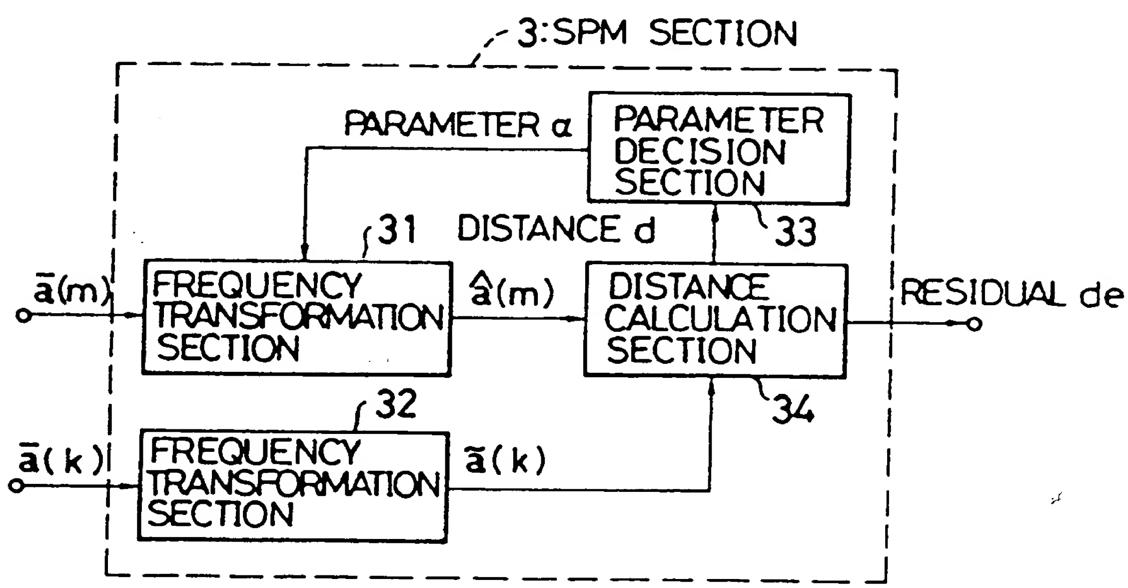

3:SPM SECTION
PARAMETER α
PARAMETER DECISION SECTION
31
DISTANCE d
33
$\bar{a}(m)$
FREQUENCY TRANSFORMATION SECTION
$\hat{a}(m)$
DISTANCE CALCULATION SECTION
RESIDUAL de
32
34
$\bar{a}(k)$
FREQUENCY TRANSFORMATION SECTION
$\tilde{a}(k)$

FIG.3A

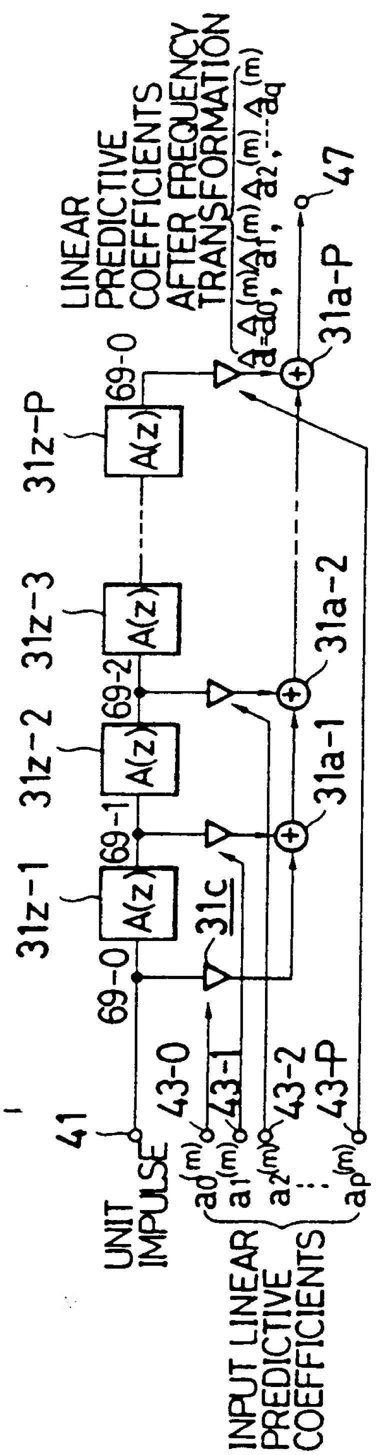

UNIT IMPULSE
41
INPUT LINEAR PREDICTIVE COEFFICIENTS
$a_0^{(m)}$
$a_1^{(m)}$
$a_2^{(m)}$
$a_p^{(m)}$
43-0
43-1
43-2
43-P
69-0
69-1
69-2
31c
31z-1
31z-2
31z-3
31z-P
A(z)
31a-1
31a-2
31a-P
47
LINEAR PREDICTIVE COEFFICIENTS AFTER FREQUENCY TRANSFORMATION
$\hat{a}=\hat{a}_0^{(m)}, \hat{a}_1^{(m)}, \hat{a}_2^{(m)}, \cdots \hat{a}_q^{(m)}$

FIG.3B

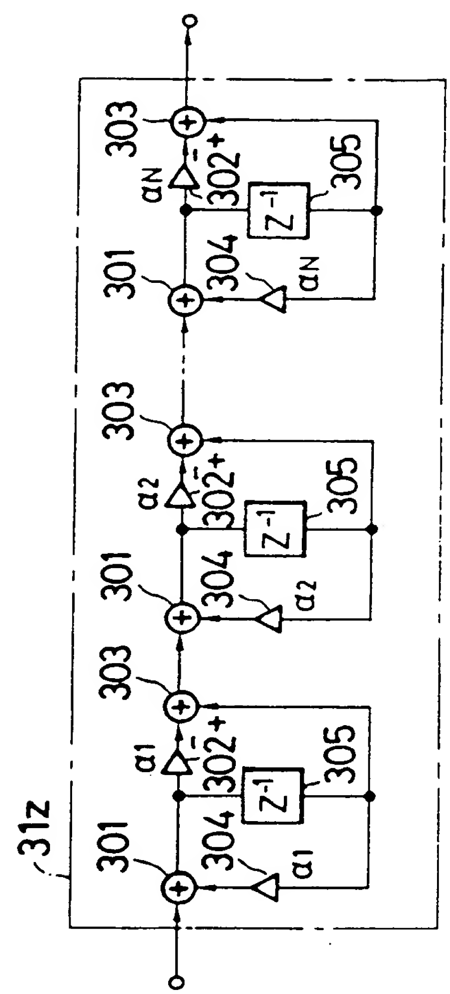

31z
301
302
303
304
305
Z^{-1}
α_1
α_2
α_N

FIG.3C

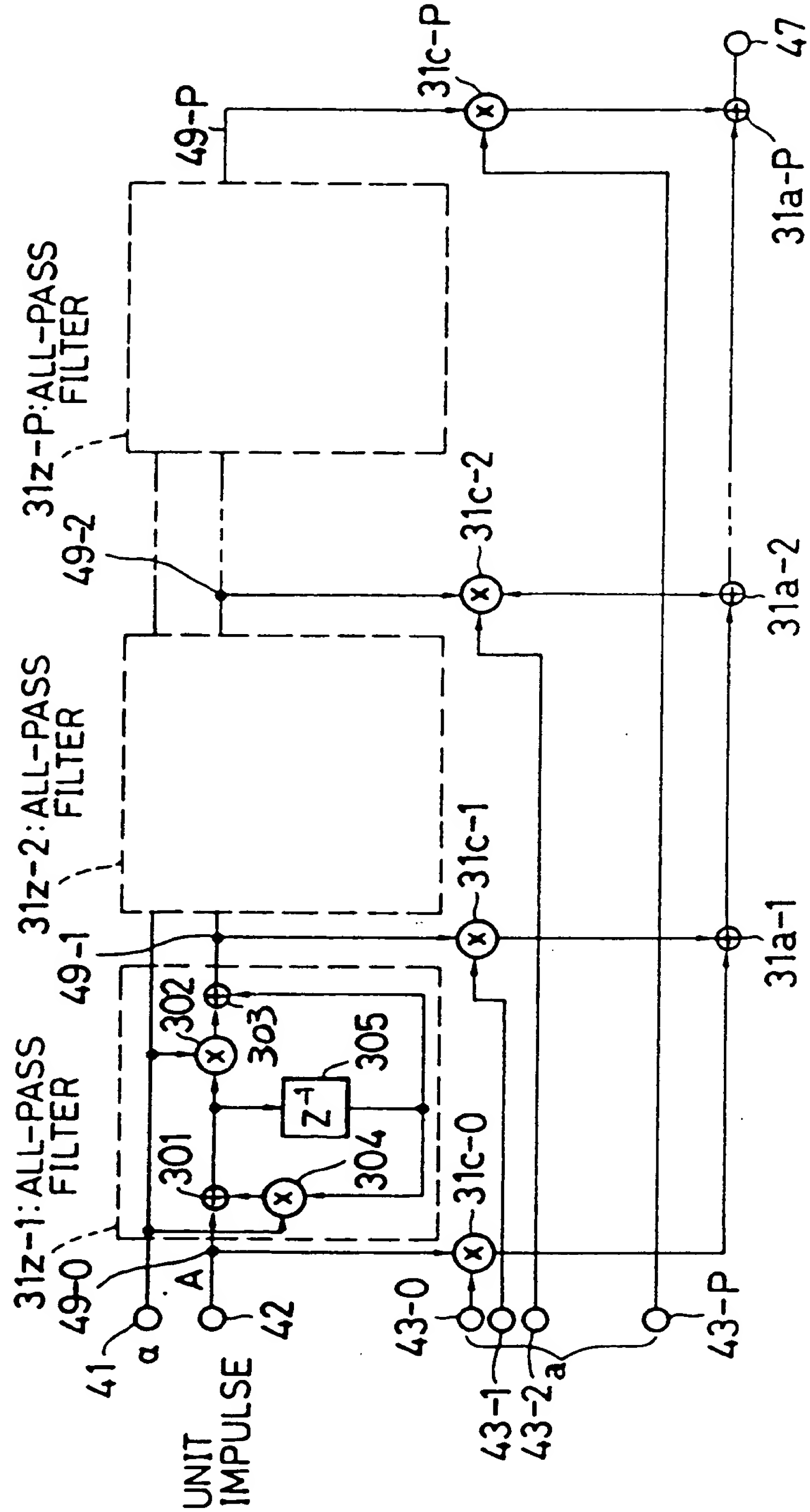

31z-1: ALL-PASS FILTER
31z-2: ALL-PASS FILTER
31z-P: ALL-PASS FILTER
49-0
49-1
49-2
49-P
301
302
303
304
305
Z-1
31c-0
31c-1
31c-2
31c-P
31a-1
31a-2
31a-P
47
41
α
42
A
UNIT IMPULSE
43-0
43-1
43-2
43-P
a

FIG.4

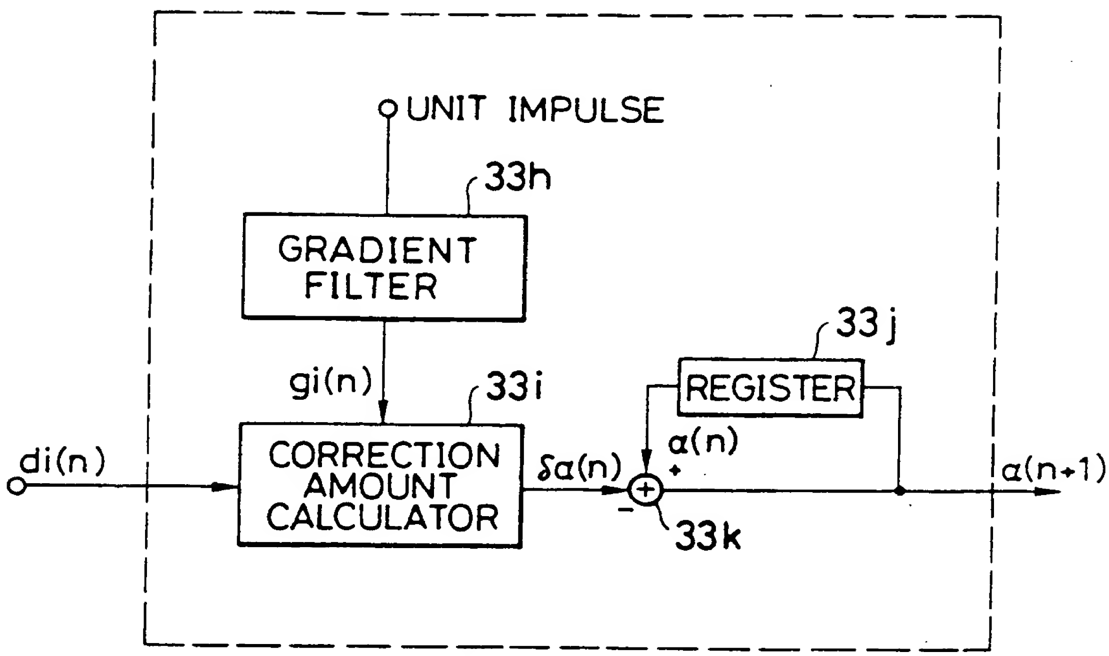

UNIT IMPULSE
33h
GRADIENT FILTER
gi(n)
33i
33j
REGISTER
di(n)
CORRECTION AMOUNT CALCULATOR
δα(n)
α(n)
+
−
33k
α(n+1)

FIG.5A

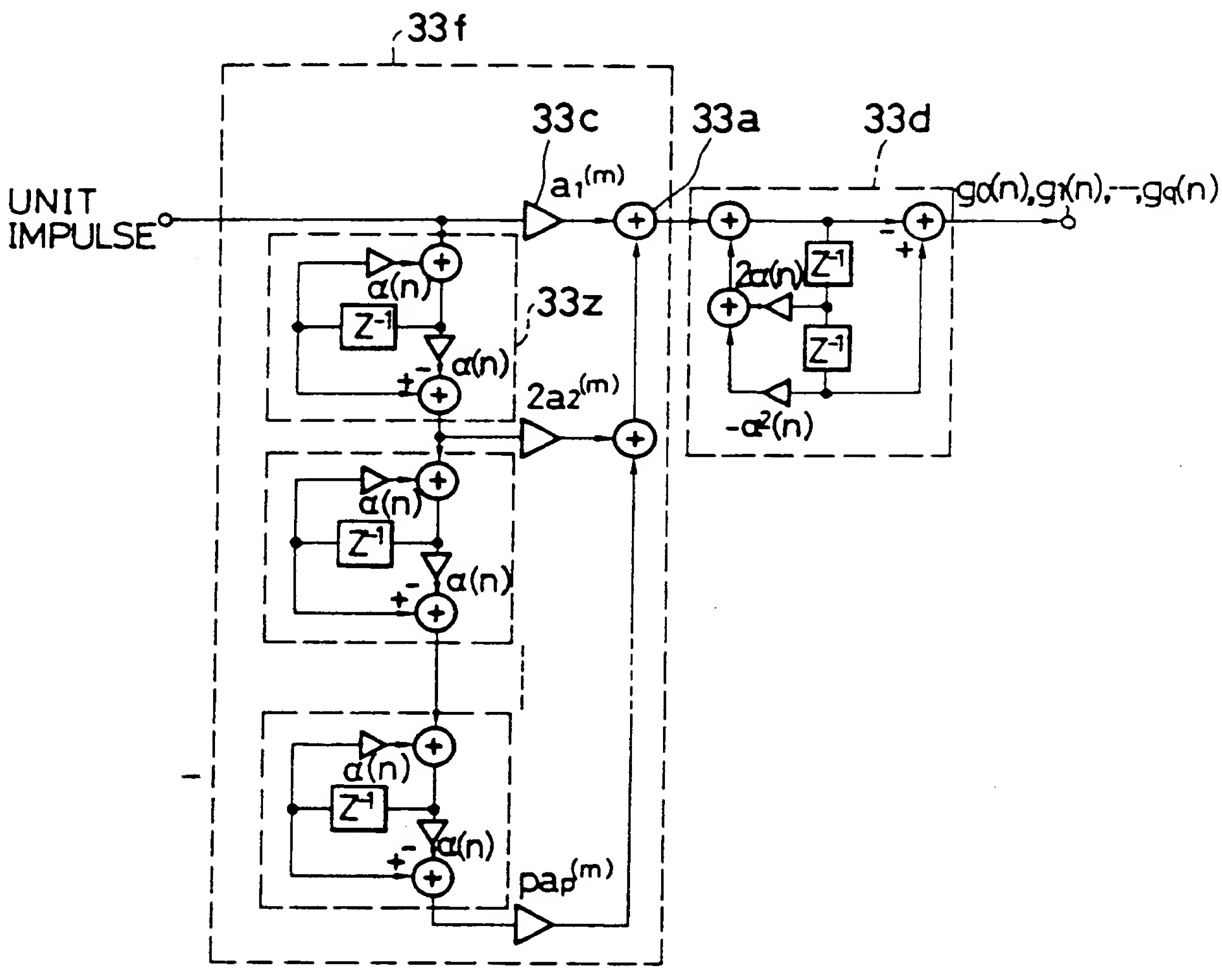

33f
33c
33a
33d
UNIT IMPULSE
$a_1^{(m)}$
$g_0(n), g_1(n), \cdots, g_q(n)$
α(n)
Z^{-1}
33z
$2a_2^{(m)}$
2α(n)
$-α^2(n)$
$pa_p^{(m)}$

FIG.5B

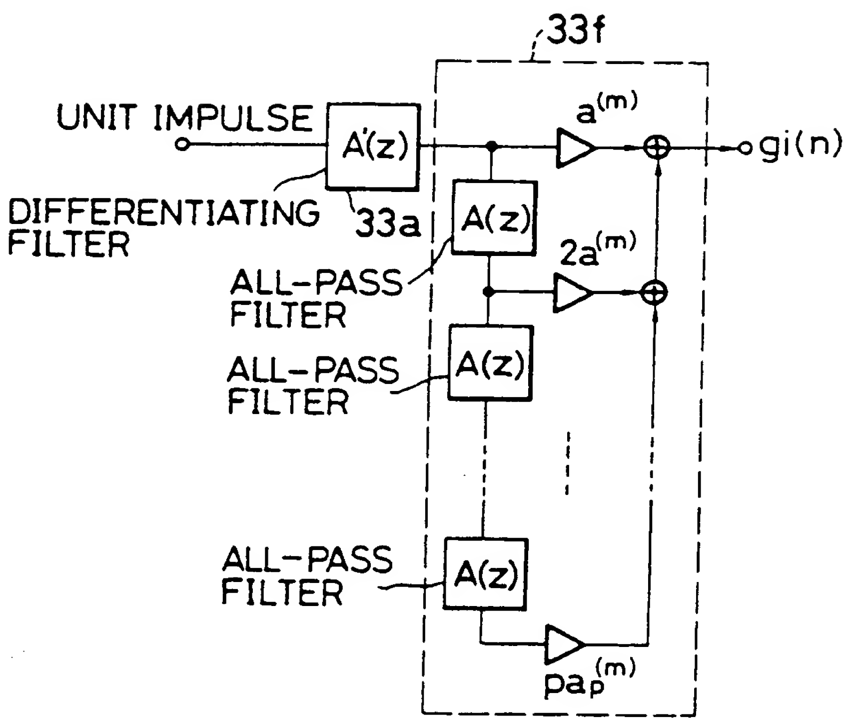

33f
UNIT IMPULSE
A'(z)
DIFFERENTIATING FILTER
33a
A(z)
ALL-PASS FILTER
A(z)
ALL-PASS FILTER
A(z)
ALL-PASS FILTER
A(z)
a(m)
2a(m)
pap(m)
gi(n)

FIG.6

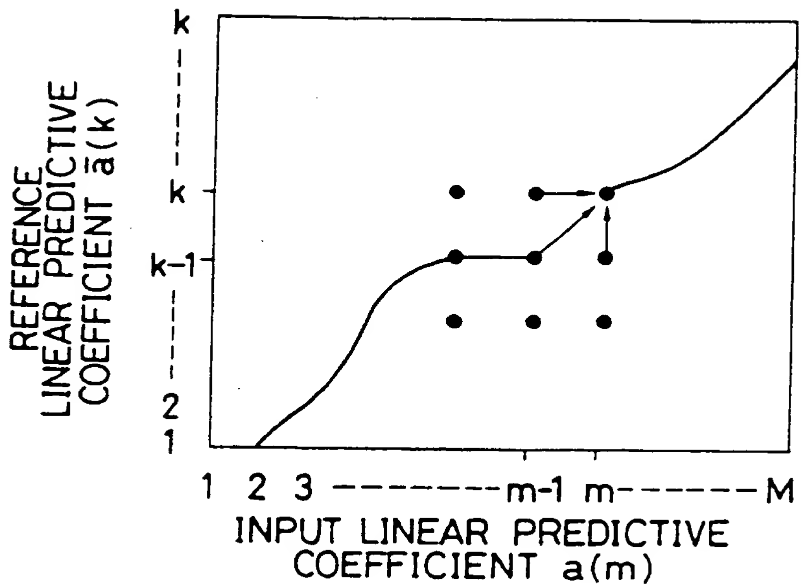

REFERENCE LINEAR PREDICTIVE COEFFICIENT ā(k)
k
k
k-1
2
1
1 2 3 m-1 m M
INPUT LINEAR PREDICTIVE COEFFICIENT a(m)

FIG.7

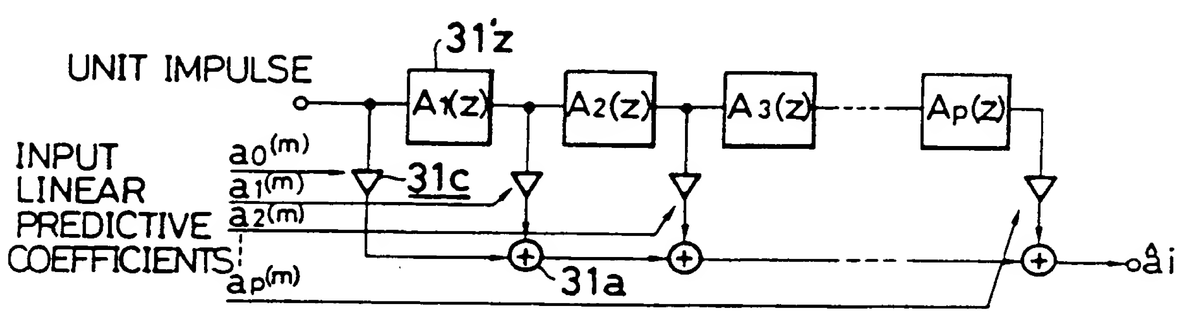
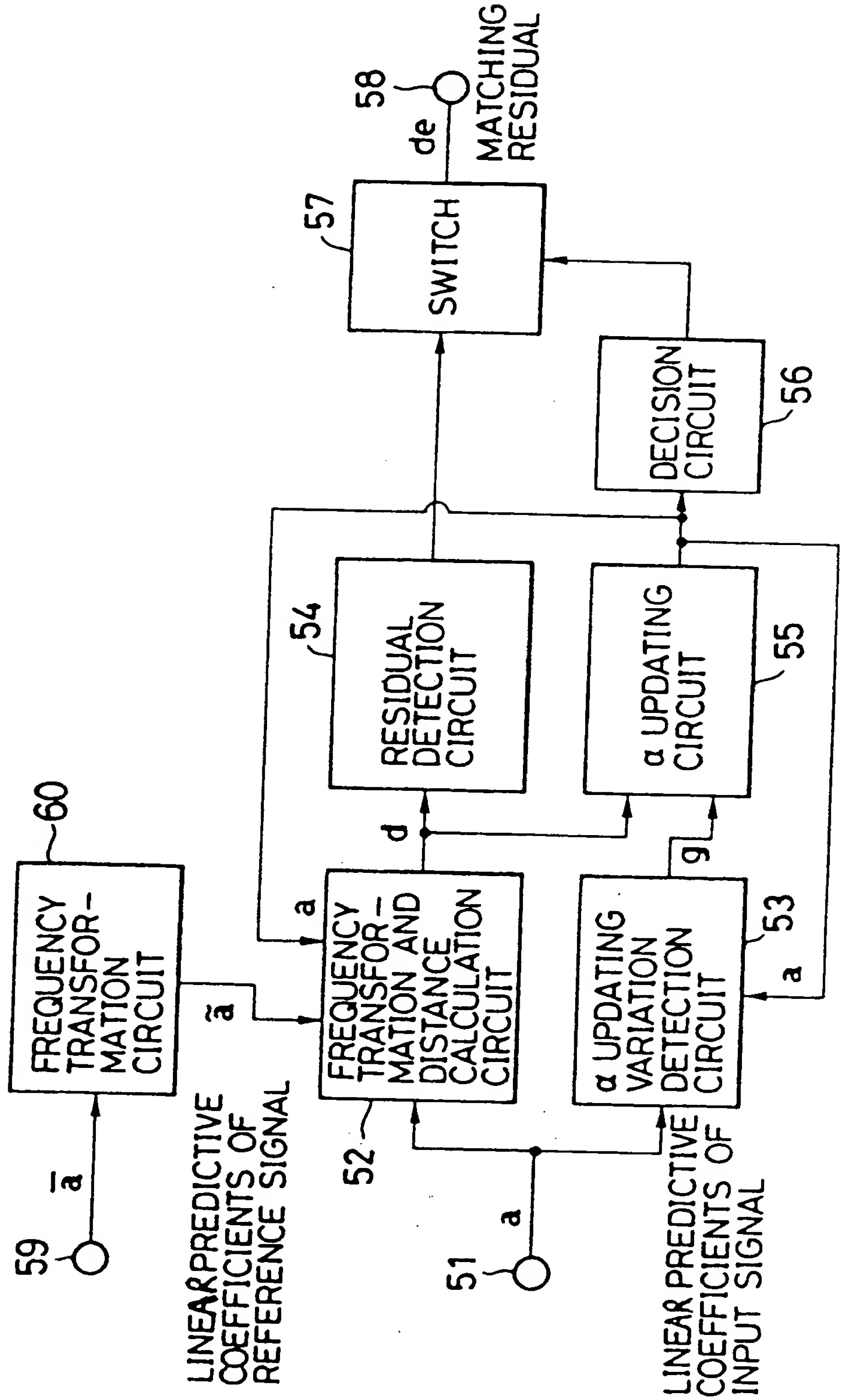

UNIT IMPULSE
31'z
A1(z)
A2(z)
A3(z)
Ap(z)
INPUT LINEAR PREDICTIVE COEFFICIENTS
a0(m)
a1(m)
a2(m)
ap(m)
31c
31a
âi

FIG.8


59
$\bar{a}$
LINEAR PREDICTIVE COEFFICIENTS OF REFERENCE SIGNAL
FREQUENCY TRANSFOR-MATION CIRCUIT
60
$\tilde{a}$
51
a
LINEAR PREDICTIVE COEFFICIENTS OF INPUT SIGNAL
52
FREQUENCY TRANSFOR-MATION AND DISTANCE CALCULATION CIRCUIT
a
α UPDATING VARIATION DETECTION CIRCUIT
53
a
d
g
RESIDUAL DETECTION CIRCUIT
54
α UPDATING CIRCUIT
55
DECISION CIRCUIT
56
SWITCH
57
de
58
MATCHING RESIDUAL

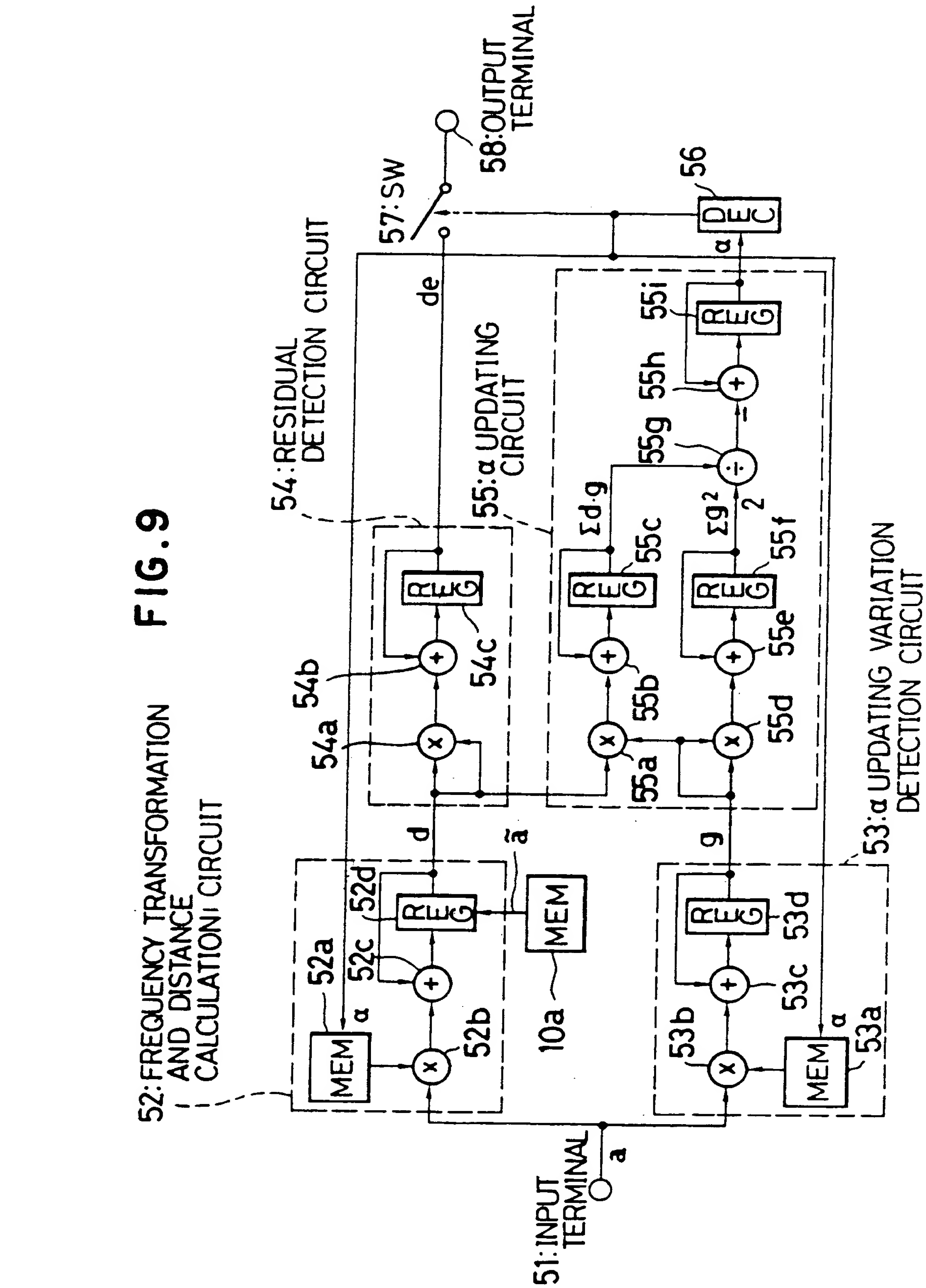

51: INPUT TERMINAL
a
52: FREQUENCY TRANSFORMATION AND DISTANCE CALCULATION CIRCUIT
MEM
52a
α
52b
52c
52d
REG
d
ã
MEM
10a
53b
53c
53d
53a
53: α UPDATING VARIATION DETECTION CIRCUIT
g
54a
54b
54c
54: RESIDUAL DETECTION CIRCUIT
55: α UPDATING CIRCUIT
55a
55b
55c
55d
55e
55f
55g
55h
55i
Σd·g
Σg²
2
de
56
DEC
57: SW
58: OUTPUT TERMINAL

FIG. 9

FIG. 10

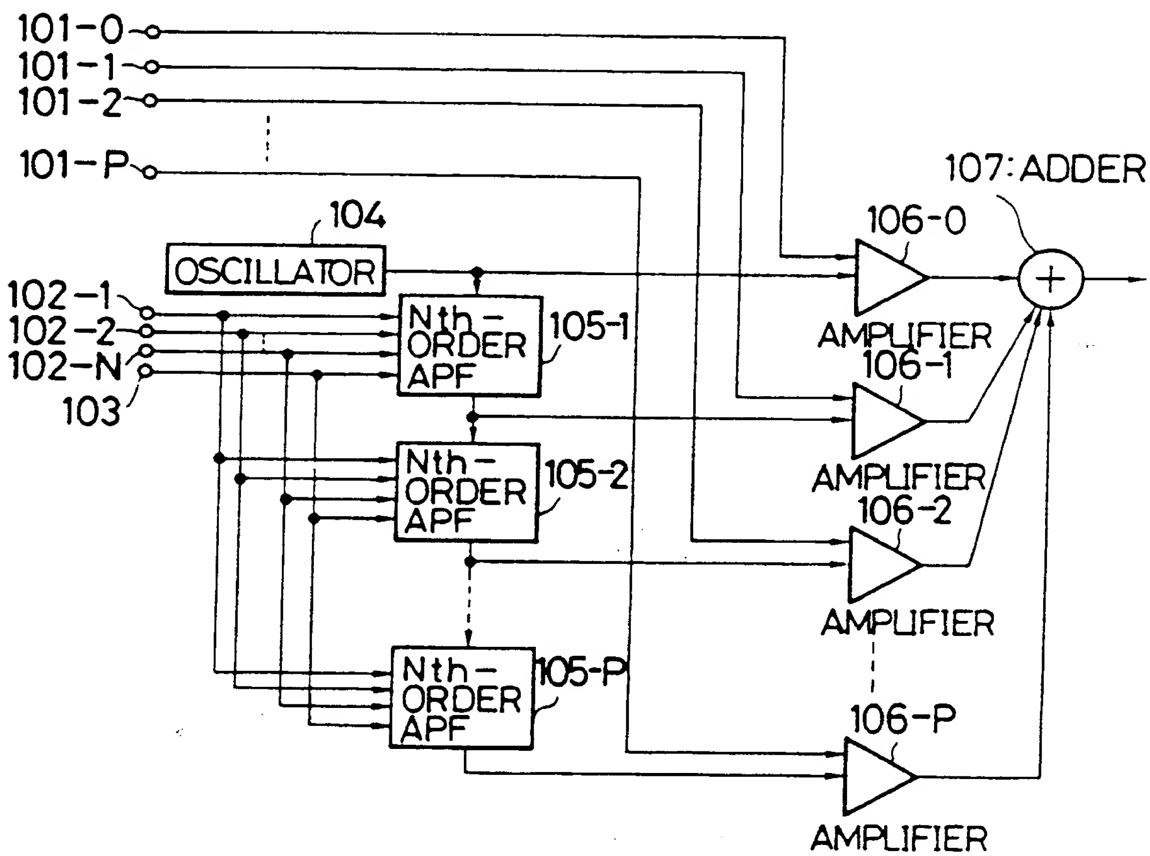

101-0
101-1
101-2
101-P
107: ADDER
104
OSCILLATOR
106-0
102-1
102-2
102-N
103
Nth-ORDER APF
105-1
AMPLIFIER
106-1
Nth-ORDER APF
105-2
AMPLIFIER
106-2
AMPLIFIER
Nth-ORDER APF
105-P
106-P
AMPLIFIER

FIG.11A

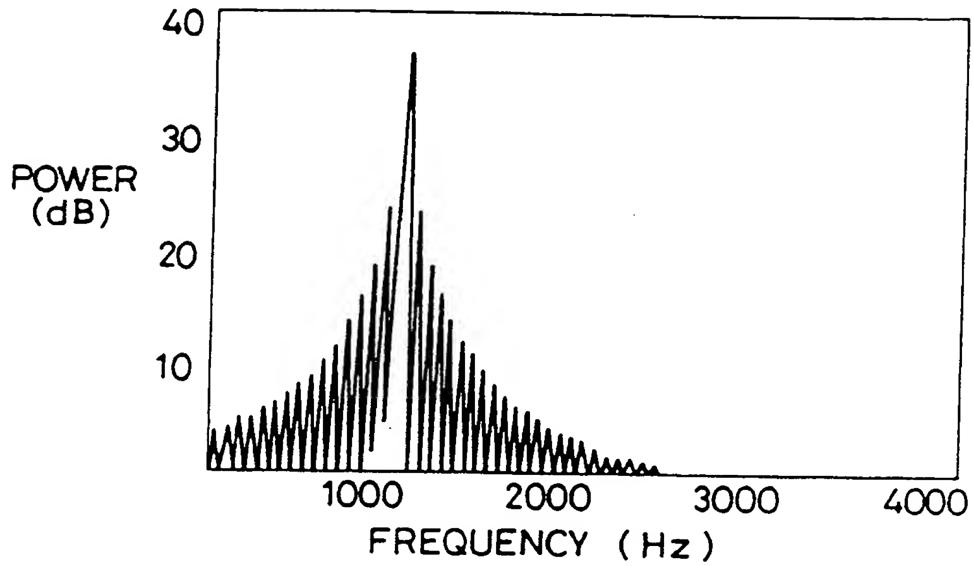


FIG.11B

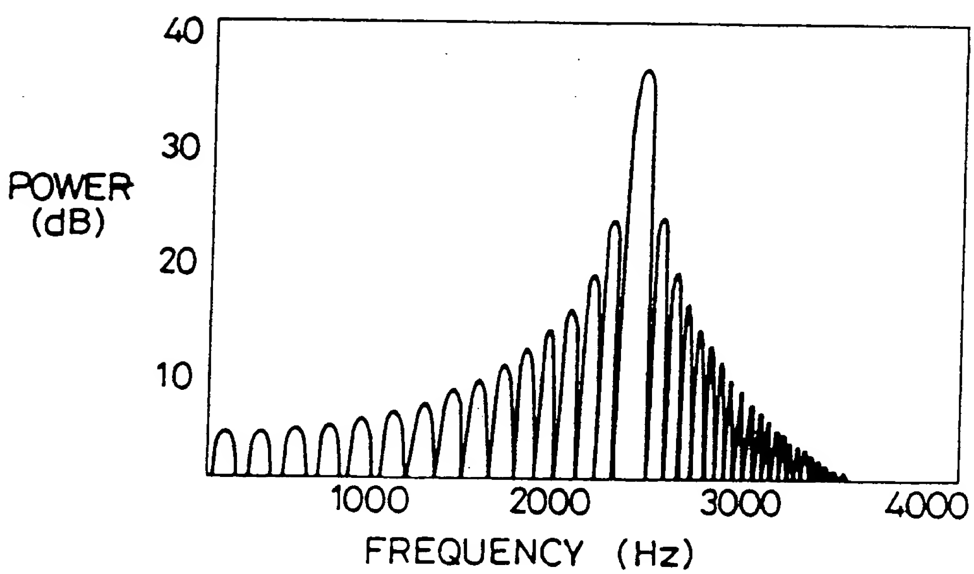


FIG.12

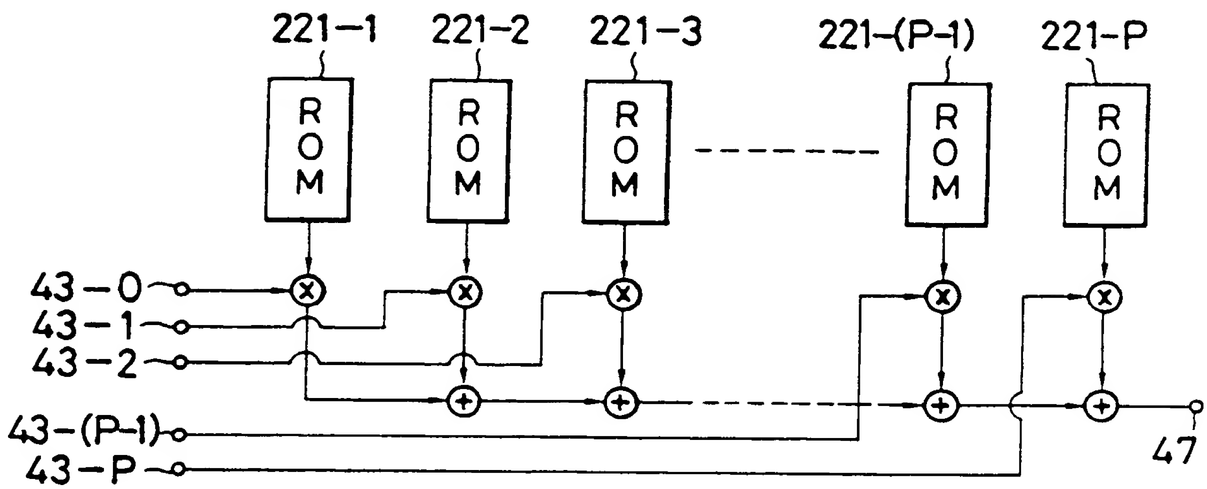
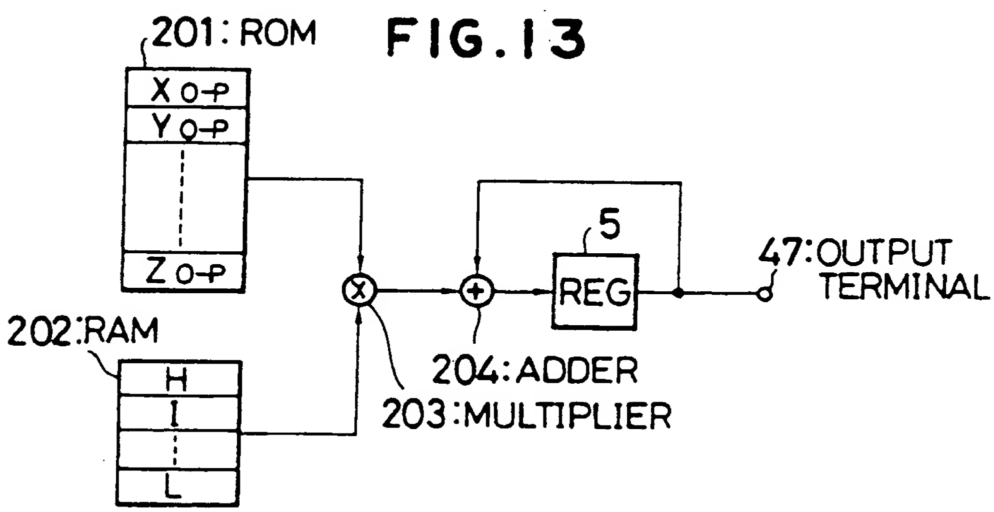

221-1
221-2
221-3
221-(P-1)
221-P
ROM
ROM
ROM
ROM
ROM
43-0
43-1
43-2
43-(P-1)
43-P
47

FIG.13


201: ROM
X 0-P
Y 0-P
Z 0-P
202: RAM
H
I
L
5
REG
47: OUTPUT TERMINAL
204: ADDER
203: MULTIPLIER